

# Testing Threats in Repeated Games

Ran Spiegler \* <sup>†</sup>

November 6, 2001

## Abstract

I introduce a solution concept for infinite-horizon games, called “Nash equilibrium with added tests”, in which players optimize with respect to relevant threats only after having tested them before. Both the optimal response and the tests are part of equilibrium behavior. The concept is applied to repeated  $2 \times 2$  games and yields the following results:

- Sustained cooperation in games such as the Prisoner’s Dilemma is preceded by a “build up” phase, whose comparative statics are characterized.
- Sustainability of long-run cooperation by means of familiar self-enforcement conventions varies with the payoff structure. E.g., “constructive reciprocity” achieves cooperation with minimal build-up time in the Prisoner’s Dilemma, yet it is inconsistent with long-run cooperation in Chicken.
- Nevertheless, a “folk theorem” holds for this class of games.

---

\*Eitan Berglas School of Economics, Tel Aviv University, Tel Aviv, Israel 69978. E-mail: [rani@post.tau.ac.il](mailto:rani@post.tau.ac.il). [Http://www.tau.ac.il/~rani](http://www.tau.ac.il/~rani).

<sup>†</sup>I undertook this project while a research fellow at Nuffield College and a member at the Institute for Advanced Study, Princeton. I am deeply grateful to both institutions for their hospitality. I owe special thanks to Eric Maskin for his fruitful suggestions concerning the constructive proof of the folk theorem. I also thank Eddie Dekel, Kfir Eliaz, Shachar Kariv, Ariel Rubinstein, Larry Samuelson, Joel Sobel, Asher Wolinsky, as well as theory seminar participants at Berkeley, Caltech, Columbia, NYU, Penn, Princeton, Stanford, UCLA, UCSD, U. of Chicago, Wisconsin and Yale, for their helpful comments.

# 1 Introduction

The theory of infinitely repeated games with perfect monitoring sustains non-myopic behavior, as part of an optimal response to threats to “punish” unwarranted myopic play. The reason that players acknowledge these threats varies across equilibrium concepts. In Nash equilibrium (NE), a player’s belief in the reality of his opponent’s threats is simply assumed to be correct. In subgame perfect equilibrium (SPE), the threats are also required to be “credible”, in the following sense: had the player tested the threats, the opponent would have wanted to carry them out.

This paper studies a different “source of credibility” of threats in repeated games: players acknowledge threats that make them behave non-myopically, simply because they **actually** tested the threats in the past and witnessed their actual realization. In other words, players optimize against a relevant threat only after having tested it. I refer to such tests as “**T-tests**”.

To illustrate this idea, consider the infinitely repeated Prisoner’s Dilemma (PD henceforth) with discounting. As usual, stage-game payoffs satisfy:  $u(C, C) > \frac{1}{2}u(D, C) > u(D, D) > u(C, D) = 0$ . Suppose that player 2 plays the strategy represented by Figure 1:

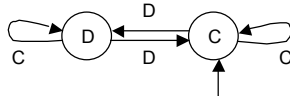


Figure 1

If player 1 is sufficiently patient, strict optimization prescribes that he play  $C$  against player 2’s  $C$ -state. This is justified by player 2’s threat to punish defection against the  $C$ -state. If the threat did not exist - i.e., if player 2 continued to play  $C$  regardless of player 1’s behavior - cooperation would cease to be optimal for player 1. Hence, the threat associated with the  $C$ -state affects player 1’s optimal response - it is “relevant”. Our model of behavior requires player 1 to test the threat (by playing  $D$ ) before he optimizes against it. In contrast, the threat associated with player 2’s  $D$ -state (to remain in that state if player 1 cooperates) prescribes the stage-game best-reply ( $D$ )

for player 1. This would also have been the optimal action in the absence of the threat. Hence, the threat is “irrelevant” and will never be tested by player 1.

Why would a player want to run T-tests? An obvious motivation is strategic uncertainty: the player is not absolutely confident of the threats he ascribes to his opponent; to verify their reality, he runs T-tests. However, there are other possible rationales for T-tests, even when the player perfectly knows his opponent’s strategy. E.g., the player may be tempted by the myopic response; he must actually experience the consequences of succumbing to the temptation before he can start serving his long-run interests. Alternatively, the player may need to justify his behavior ex-post to a principal and realizing the opponent’s threats may be necessary in order to persuade the principal.<sup>1</sup>

At any rate, the present paper does not adhere to any particular rationale for T-tests, but simply assumes that players optimize against relevant threats only after having tested them. The question of rationalizing T-tests is deferred to the concluding section.

So far, we have dealt with the behavior of an individual player. In order to analyze repeated  $2 \times 2$  games, in which both players behave in this way, I introduce an equilibrium concept, which incorporates T-tests into its very definition. A “**Nash equilibrium with added tests**” (**NEWT**) is a pair of pure strategies that induce a play path, in which players optimize against relevant threats only after having tested them finitely many times; the T-tests come to an end in finite time and the players eventually adhere to best-replying.

In NEWT, each player follows a single strategy from beginning to end, which carries out both the best-replying and the T-tests with respect to the opponent’s strategy. Thus, NEWT is **not** a refinement of NE because it allows for sub-optimal actions (the T-tests) along the play path. For the sake of methodological parsimony, NEWT accommodates no further departures from best-replying, apart from T-tests.

Figure 2 presents an example of a NEWT in the repeated PD, in finite automata representation. Both  $s_1$  and  $s_2$  contain two relevant threats, associated with the states  $q^0, q^2$  and  $r^1, r^2$ . Player 1 (2) departs from best-replying at periods 2 and 3 (1 and 4). All of these departures are T-tests that back-up

---

<sup>1</sup>See Spiegler (2000) for a formalization of this idea, and Tetlock and Boettger (1989,1994) for experimental evidence on the justifiability motive.

subsequent non-myopic optimal responses to the relevant threats.

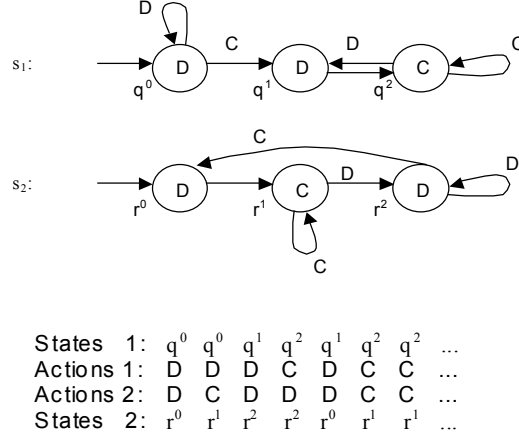


Figure 2

Our analysis of NEWT contains two novel contributions to the study of long-term cooperation: (1) characterization of a “build-up” phase that precedes cooperation in repeated  $2 \times 2$  games; (2) re-examination of the sustainability of long-run cooperation by means of familiar self-enforcement conventions, such as reciprocity or trigger strategies.

The NEWT given by Figure 2 exhibits two interesting features. First, players move gradually from mutual defection to mutual cooperation, through a five-period “build-up” phase. Second, equilibrium strategies display a “constructive reciprocity” property: when players are in a “cooperative” state, they continue to cooperate if and only if the opponent plays cooperatively. Both features recall similar behavior patterns reported by Selten et. al. (1997) in their experimental study of repeated duopoly.

While NE or SPE are silent over the existence and properties of a cooperation build-up phase, such a phase is a prerequisite for long-run cooperation under NEWT. Moreover, the build-up phase of Figure 2 is the **shortest** possible in the repeated PD. The minimal cooperation build-up length increases as  $u(C, C)$  decreases and the players become more patient. The minimal

build up can always be achieved in NEWT by strategies that display constructive reciprocity.

Surprisingly, constructive reciprocity turns out to be **inconsistent** with sustained cooperation in the repeated game of “**Chicken**”. Thus, while constructive reciprocity is an efficient convention for sustaining cooperation in repeated PD under NEWT, it obstructs the attainment of the same goal in repeated Chicken. By comparison, mutual cooperation is NE- and SPE-sustainable in both PD and Chicken, by strategies that display constructive reciprocity. This result reveals a sense, in which “developing cooperation in collective action problems” and “attaining peace in conflicts over a scarce resource” are two inherently different problems.

Results of similar flavor are obtained for “trigger” strategies, perhaps the simplest and most familiar self-enforcement convention in repeated games. “Trigger” strategy structures, both “grim” and “forgiving”, that are sufficient to generate Nash and perfect “folk theorems” in repeated symmetric  $2 \times 2$  games, turn out to be highly restrictive under NEWT.

If the performance of these simple self-enforcement conventions is so sensitive to the payoff structure, the question arises, whether a NEWT folk theorem holds for repeated  $2 \times 2$  games. The answer turns out to be affirmative: every individually rational payoff profile in repeated  $2 \times 2$  games can be approximated by some NEWT. It is unclear whether this result is extendible to larger games.

The paper proceeds as follows. Section 2 defines the concept of NEWT for simultaneous-move, repeated  $2 \times 2$  games. Section 3 uses NEWT to revisit the problem of sustaining cooperation in repeated PD and Chicken. Section 4 analyzes NEWT with trigger strategies in repeated symmetric  $2 \times 2$  games. Section 5 presents the folk theorem. Section 6 extends the concept of NEWT to a broader class of infinite-horizon games and analyzes a few examples (including a repeated Ultimatum game). Section 7 discusses the interpretation of the equilibrium concept. Some proofs are relegated to the appendix.

## 2 The Equilibrium Concept

Two players play an infinitely repeated, simultaneous-moves  $2 \times 2$  games with discounting. Denote the action set (for both players) by  $A$ . Player  $j$ ’s stage-game payoff function is denoted by  $u_j$  and his opponent’s identity is denoted

by  $-j$ . For every  $a_j \in A$ , denote  $-j$ 's stage-game ("myopic") best-reply by  $br_{-j}(a) \in A$ . Stage-game indifference are assumed away. For every  $a \in A$ , denote  $-a = A \setminus \{a\}$ .

We will only consider pure strategies that admit a finite automata representation  $(Q_j, q_j^0, f_j, \tau_j)$ , where:  $Q_j$  is a finite set of states;  $q_j^0$  is the initial state;  $f_j : Q_j \rightarrow A$  is an output function, which specifies the action taken by player  $j$  when he is in state  $q \in Q_j$ ; and  $\tau_j : Q_j \times A \rightarrow Q_j$  is a transition function, which specifies the state to which the automaton switches from state  $q \in Q_j$  when the opponent plays  $a_{-j} \in A$  against  $q$ . The transition from  $q$  is said to be constant if  $\tau_j(q, a) = \tau_j(q, -a)$ .<sup>2</sup> Note that a pure strategy has an infinite number of finite automata representation. We will always identify a strategy  $s_j$  with a fixed representation  $(Q_j, q_j^0, f_j, \tau_j)$ .

Let  $z(s_1, s_2) = ((a_1^k, a_2^k))_{k=0,1,2,\dots}$  be the play path induced by the strategy profile  $(s_1, s_2)$ , where  $a_j^k \in A$  is player  $j$ 's action at period  $k$ . I.e.,  $a_j^0 = f_j(q_j^0)$ ,  $a_j^1 = f(\tau_j(q_j^0, a_{-j}^0))$ , and so forth. Given  $z(s_1, s_2)$ , the function  $p_j : \{0, 1, 2, \dots\} \rightarrow Q_j$  keeps track of the state of player  $j$ 's strategy at period  $k$  along  $z(s_1, s_2)$ . Using Figure 2 for illustration,  $BR_2(q^0) = BR_2(q^2) = C$  and  $BR_2(q^1) = D$ ; along the play path,  $p_1(0) = p_1(1) = q^0$ ,  $p_1(2) = q^1$ , etc.

Because preferences satisfy the discounting criterion,  $s_j$  induces a well-defined correspondence  $BR_{-j} : Q_j \rightarrow A$ , which assigns to every  $q \in Q_j$  the set of actions for  $-j$  that are consistent with best-replying to  $s_j$ , at any period  $k$  for which  $p_j(k) = q$ . For expositional simplicity only, assume that  $BR_{-j}(q)$  is a **singleton** for every  $q \in Q_j$ . It is referred to as  $-j$ 's "long-run best-reply" action against  $q$ . For example, in Figure 1,  $BR(q^C) = C$  and  $BR(q^D) = D$ . Whenever  $\tau_j(q, \cdot)$  is constant,  $BR_{-j}(q) = br_{-j}[f_j(q)]$ .

Given  $(s_1, s_2)$ , let  $e(q)$  be the number of times  $BR_{-j}(q)$  is played against  $q \in Q_j$ , before  $BR_{-j}(q)$  is played against  $q$  **for the first time** along  $z(s_1, s_2)$ . If  $BR_{-j}(q)$  is never played against  $q$  along  $z(s_1, s_2)$ , then  $e(q)$  is simply the total number of periods along  $z(s_1, s_2)$ , for which  $p_j(k) = q$ . E.g., in Figure 2,  $e(q^0) = e(q^2) = 1$ ,  $e(q^1) = 0$ .

We are ready to formulate an equilibrium concept for repeated  $2 \times 2$  games, which reflects the idea that players respond optimally to relevant threats only after having tested them.

---

<sup>2</sup>The finite automaton given in Figure 1 is:  $Q = \{q^C, q^D\}$ ;  $q^0 = q^C$ ;  $f(q^C) = C$  and  $f(q^D) = D$ ;  $\tau(q, C) = q$ ; and  $\tau(q, D) = Q \setminus \{q\}$  for every  $q \in Q$ .

**Definition 1**  $(s_1, s_2)$  is a *Nash equilibrium with added tests (NEWT)* if for every player  $j = 1, 2$ :

1.  $a_{-j}^k = BR_{-j}[p_j(k)]$  for every sufficiently large period  $k$  along  $z(s_1, s_2)$ .
2.  $e(q) > 0$  if and only if  $BR_{-j}(q) \neq br_{-j}[f_j(q)]$ , for every state  $q \in Q_j$  that is visited along  $z(s_1, s_2)$ .

The first condition says that eventually, players stick to best-replying. The second condition says that they test threats that affect optimal response, finitely many times, before optimizing against them. A non-myopic action is optimal only in the face of a threat to punish the sub-optimal action, whereas the myopic action is justified even in the absence of such a threat. Therefore, testing threats occurs in the former case but not in the latter. In order for us to be able to isolate the effect of T-tests on repeated-game behavior, Definition 1 allows no further departures from best-replying.

It is known that a path induced by finite automata eventually enters a cycle (see Osborne and Rubinstein (1994, Ch. 8)). I.e., there exist an earliest period  $k^*$  and an integer  $L$ , such that  $p_j(k) = p_j(k + L)$  for every  $j = 1, 2$ ,  $k \geq k^*$ . Thus, all T-tests take place prior to  $k^*$ . Starting at period  $k^*$ , players play a repeated-game NE  $(s'_1, s'_2)$ , where  $s'_j$  is identical to  $s_j$ , except that the initial state is  $p_j(k^*)$  instead of  $p_j(1) = q_j^0$ . NEWT can be viewed as a NE with altered initial states, such that the  $e(q)$ 's obey Definition 1.

The incorporation of T-tests into a notion of equilibrium behavior raises interesting questions of interpretation, which will be discussed at length in **Section 7**. In the meantime, let us conclude the present section with a number of comments.

1. Definition 1 requires T-tests against any relevant threat that is encountered along the **entire** play path, not just those that are encountered in its cyclic phase. E.g., in Figure 2, the threat associated with state  $r^2$  is encountered only in the pre-cyclic phase. It would not have been encountered if player 1 had not run the T-test against  $r^1$ . Nevertheless, player 1 runs a T-test against  $r^2$ .
2. There are three possible kinds of violation of NEWT:

- (a) The cyclic phase of the play path contains sub-optimal behavior.
  - (b)  $BR_j[p_{-j}(k)] = br_j(a_{-j}^k)$  and  $a_j^k \neq BR_j[p_{-j}(k)]$  for some  $k$ .
  - (c)  $a_j^k = BR_j[p_{-j}(k)] \neq br_j(a_{-j}^k)$  and there is no  $l < k$  satisfying  $p_{-j}(l) = p_{-j}(k)$ ,  $a_j^l = br_j(a_{-j}^l)$ .
3. Every non-zero-sum  $2 \times 2$  game has a pure-strategy NE. An infinite repetition of this NE is consistent with NEWT because it displays both long-run and myopic best-replying, so no T-tests are needed. It follows that a NEWT exists for every repeated non-zero-sum  $2 \times 2$  game.
4. In general, every repeated-game NE, whose induced (eventually cyclic) play path consists of nothing but stage-game Nash equilibria, is also a NEWT. Every other NE contains non-myopic behavior that is sustained by untested threats and is therefore not a NEWT. Likewise, every other NEWT contains sub-optimal behavior and is therefore not a NE. Thus, the intersection between the class of NEWT and the class of NE consists of the NE, in which players play a stage-game NE at every period.

Of course, we are mostly interested in characterizing the class of non-NE NEWT. It is to this task that we now turn.

### 3 Building up Cooperation in PD and Chicken

We have already seen in Section 1 that sustained cooperation in repeated PD is NEWT-sustainable. In Figure 2, players start cooperating indefinitely after a build-up phase that lasts **five** periods. This turns out to be the shortest build-up phase under NEWT. Formally, let  $k^C$  be the earliest period, for which there exists a NEWT in the repeated PD, such that along the equilibrium path,  $a_1^k = a_2^k = C$  for every  $k \geq k^C$ .

**Theorem 1**  $k^C \geq 6$ . Moreover,  $k^*$  increases as players become more patient and as  $u(C, C)$  decreases.

The comparative statics are intuitive. The lower  $u(C, C)$  or the more patient players are, the harsher the punishment that is needed to sustain cooperation. Since this punishment must be realized along the equilibrium path, the build-up phase is longer.



The strategies used in the construction of the minimal build-up display a familiar convention for self-enforcing cooperation, to which I shall refer as “**constructive reciprocity**”.<sup>3</sup> When players play cooperatively at period  $t$ , they reciprocate - i.e., they cooperate at  $t + 1$  if and only if the opponent cooperates at  $t$ . (See Figures 1, 2 and 8). Formally, let  $a^* = \arg \min_{a \in A_2} [\max_{b \in A_1} u_1(b, a)]$  be the minimax action in a symmetric  $2 \times 2$  game. (the definition for asymmetric games is straightforward.)

**Definition 2** *A strategy  $s = (Q, q^0, f, \tau)$  in a repeated symmetric  $2 \times 2$  game satisfies **constructive reciprocity**, if  $f[\tau(q, a)] = a$  for every  $a \in A$  and every  $q \in Q$  satisfying  $f(q) = -a^*$ .*

Such strategies are also referred to as **CR strategies**. No restrictions are imposed on transitions from  $a^*$ -states or on the initial state.

Mutually cooperation is NE- and SPE-sustainable by CR strategies, not only in repeated PD but also in repeated Chicken. As CR strategies sustain cooperation in repeated PD in the most efficient way possible under NEWT, the question arises whether this property holds for the game of “Chicken”, given by the following payoff matrix. As usual, assume  $b > a > \frac{b+1}{2} > 1$ . The minimax action is “Hawk”.

	<i>Dove</i>	<i>Hawk</i>
<i>Dove</i>	$a, a$	$1, b$
<i>Hawk</i>	$b, 1$	$0, 0$

**Theorem 2** *There exists no NEWT with CR strategies in repeated Chicken, in which  $a_1^k = a_2^k = \text{“Dove”}$  for every sufficiently large  $k$ .*

**Proof.** Let  $D$  and  $H$  stand for “dove” and “hawk”. First, observe that by constructive reciprocity and the payoff structure of the game,  $BR(q) = D$  whenever  $f(q) = D$ . Thus,  $f[\tau(q, D)] = D$  implies  $BR(q) = D$  for every  $q$ .

---

<sup>3</sup>This term is due to Joel Sobel.

Denote the earliest period, in which both players play  $D$ , by  $t^*$ . By the structure of CR strategies,  $a_1^t = a_2^t = D$  for every  $t > t^*$ . Therefore, by NEWT, each player plays  $H$  at least once prior to  $t^*$ . If  $a_j^{t^*-1} = D$  and  $a_{-j}^{t^*-1} = H$ , then  $s_j$  violates constructive reciprocity. Therefore, it must be that  $a_1^{t^*-1} = a_2^{t^*-1} = H$ .

Let  $k^*$  denote the earliest period  $k$ , in which  $a_1^k = a_2^k = H$ . By NEWT,  $BR_{-j}[p_j(k^*)] = H$  for both  $j = 1, 2$ , because  $H \neq br(H)$ . By NEWT, for each player  $j = 1, 2$ , there exists  $k_j < k^*$ , such that  $p_j(k_j) = p_j(k^*)$  and  $a_{-j}^{k_j} = D$ . Let  $k_1 < k_2$ , without loss of generality. By the structure of CR strategies,  $a_2^{k_1+1} = H$ . By the definition of  $k^*$ ,  $a_1^{k_1+1} = D$ . It follows that  $f_1[\tau_1(p_1(k^*), D)] = D$ . Therefore,  $BR_2[(p_1(k^*))] = D$ , a contradiction. ■

Thus, the problem of sustaining cooperation with CR strategies is radically different for PD and for Chicken. The key to this result is that under NEWT with CR strategies, “peace” (mutually dovish play) must be preceded by at least one “war” (mutually hawkish play). Since  $H$  is not a myopic best-reply to itself, this must be backed up by T-tests as well, but these are accompanied by yet another “war”, which triggers further T-tests, and so forth. In contrast, mutual defection in PD is a stage-game NE, which triggers no further T-tests.

I believe that these results capture a genuine distinction, which escapes standard equilibrium concepts, regarding the real-life interactions for which PD and Chicken serve as parables. To put it figuratively, constructive reciprocity is an efficient convention for building up cooperation in “collective action” problems, but a destructive convention for overcoming “conflicts over a scarce resource”.

When constructive reciprocity is violated by at least one of the players, it is quite easy to sustain cooperation in repeated Chicken. The following play path is NEWT-sustainable:

<i>Player 1</i>	$H$	$D$	$H$	$D$	$D$	$D$	...
<i>Player 2</i>	$D$	$H$	$D$	$D$	$D$	$D$	...

The equilibrium strategies are given by Figure 3:

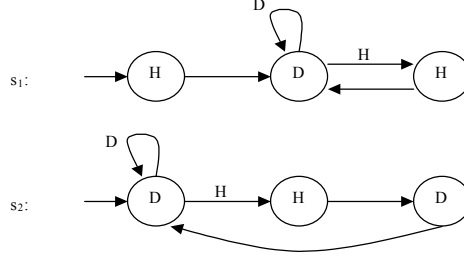


Figure 3

Note that  $s_2$  (unlike  $s_1$ ) is not a CR strategy - it does not retaliate against player 1's hawkish move at period 3. The cooperation build-up phase is the shortest possible under NEWT. The comparative statics are the same as in Theorem 1: the lower  $u(D, D)$  and the higher the discount factor, the longer the minimal build-up.

## 4 NEWT with Trigger Strategies

This section studies the performance of “trigger strategies” under NEWT. This familiar self-enforcement convention, which is rich enough to generate Nash and perfect “folk theorems” for repeated symmetric  $2 \times 2$  games, turns out to be highly restrictive under NEWT.

Let us begin with “grim” trigger strategies, which punish a deviant player by “minimizing” him for the rest of the game. Formally, a strategy  $s_j$  with a finite-automata representation  $(Q_j, q_j^0, f_j, \tau_j)$  is a **“grim” trigger strategy** if for every  $q \in Q_j$  with non-constant  $\tau_j(q, \cdot)$ ,  $f_j[o(q)] = a^*$  and  $\tau_j[o(q), \cdot] = o(q)$ , where  $o(q) = \tau_j(q, -BR_{-j}(q))$ .

In the present context, grim strategies are quite uninteresting. Since every sub-optimal action triggers a grim punishment, we should not expect non-myopic behavior to survive through the T-tests. This is indeed the case (the proof is simple and therefore omitted):

**Remark 1** In any NEWT with grim strategies,  $a_j^k = br_j(a_{-j}^k)$  for every  $j = 1, 2$  and every sufficiently large  $k$ .

Let us now consider a more interesting class of trigger strategies, which exhibits a “forgivingness” property. A state  $q \in Q$  is said to be **reachable** from  $q' \in Q$  if there exists a sequence of states  $q_0, \dots, q_m \in Q$  ( $m > 0$ ), such that  $q_0 = q'$ ,  $q_m = q$ , and for every  $k = 0, \dots, m - 1$ , there exists  $a \in A$ , such that  $\tau(q_k, a) = q_{k+1}$ .

**Definition 3** A strategy with finite automata representation  $(Q_j, q_j^0, f_j, \tau_j)$  in a repeated symmetric  $2 \times 2$  game is a **“prison” trigger strategy** if:

1. Every state  $q \in Q_j$  is reachable from any other  $q' \in Q_j$ .
2. For every  $q \in Q_j$  with non-constant  $\tau_j(q, \cdot)$ , playing  $-BR_{-j}(q)$  against  $q$  is punished by  $m(q) \geq 0$  periods, in which player  $j$  plays  $a^*$  before returning to  $q$ .<sup>4</sup>

Prison strategies consist of “normal” states and “punishment” states (these categories are not mutually exclusive). When player  $-j$  plays sub-optimally against a normal state  $q \in Q_j$  with non-constant  $\tau_j(q, \cdot)$  is “minimized” for a non-negative number  $m(q)$  of periods, which the strategy spends in punishment states, and at the end of which it returns to  $q$  (the same “situation” in which the original sub-optimal play took place). Player  $-j$ ’s behavior during the punishment phase does not affect the duration of the punishment, but this duration can vary with  $q$ . Finally, prison strategies contain no irreversibility - every state can be reached from any other state. Thus, the strategies represented by Figures 1 and 2 are not prison strategies.

Restricting attention to prison strategies is innocuous, as far as NE is concerned. Every individually rational (and eventually cyclic) play path in a

---

<sup>4</sup>More formally, whenever  $\tau_j(q, \cdot)$  is non-constant,  $Q_j$  contains a sequence of states  $(q_0, \dots, q_{m(q)+1})$ , such that:

- (a)  $q_0 = q_{m(q)+1} = q$
- (b)  $q_1 = \tau_j[q, -BR_{-j}(q)]$
- (c)  $f_j(q_n) = a^*$  and  $q_n = \tau_j(q_{n-1}, \cdot)$  for every  $n \neq 0, m(q)$ .

repeated symmetric  $2 \times 2$  game is sustainable by NE with prison strategies. This is accomplished by fixing an arbitrarily high  $m(q)$  wherever necessary. If we redefine the finite automata representation to allow players to condition on their own past moves, prison strategies also generate an SPE folk theorem for repeated symmetric  $2 \times 2$  games.

Let us turn to characterizing the class of NEWT with prison strategies in repeated symmetric  $2 \times 2$  games. Figure 4 presents a profile of prison strategies in a repeated coordination game, whose payoff structure satisfies  $u(G, G) > 3 \cdot u(B, B)$  and  $u(G, B) = u(B, G) = 0$ , where  $\{G, B\}$  is the action set. For sufficiently patient players, this strategy profile constitutes a NEWT. The cyclic phase begins at period 4. Players 1 and 2 perform T-tests at periods 3 and 2, respectively, to back-up their optimal miscoordination at period 4.

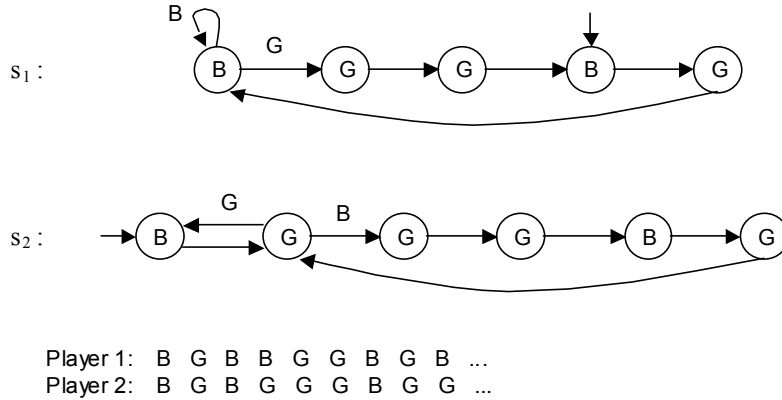


Figure 4

In contrast to NE or SPE, the restriction to prison trigger strategies turns out to be highly restrictive under NEWT.

**Theorem 3** *The only repeated symmetric  $2 \times 2$  games, for which there exists NEWT with prison strategies that contain non-myopic play, are coordination games with a Pareto-dominant outcome. Moreover, in this class of games, NEWT with prison strategies allows at most one period of miscoordination per cycle.*

The proof will make use of the following lemma (see appendix for the proof) :

**Lemma 1** *Suppose that  $(s_1, s_2)$  is a NEWT with prison strategies in a repeated symmetric  $2 \times 2$  game. If  $a_{-j}^k \neq br_{-j}(a_j^k)$  for some player  $j$ , then period  $k$  belongs to the cyclic phase of  $z(s_1, s_2)$ .*

**Proof of the theorem.** Let us first consider play paths, in which  $a_j^k \neq br_j(a_{-j}^k)$  for both players  $j = 1, 2$ , for some period  $k$ . By NEWT, there exist periods  $l_1, l_2 < k$ , such that  $p_j(l_j) = p_j(k)$  and  $a_{-j}^{l_j} = br_{-j}(a_j)$ . Consider the latest such  $l_1, l_2$  and let  $l_1 < l_2$ , without loss of generality. Then, the play pattern in the periods that immediately precede  $k$  is given as follows (note that  $-a_j = br(a_{-j})$ ):

Period	$l_1$	...	$l_2$	...	$k$
Player 1	$a_1$	...	$-a_1$	...	$a_1$
Player 2	$-a_2$	...	$a_2$	...	$a_2$

By the definition of  $k$  and the structure of prison strategies,  $a_1^m = a_2^n = a^*$  for every  $m = l_1 + 1, \dots, k - 1$  and  $n = l_2 + 1, \dots, k - 1$ . It follows that  $-a_1 = a^*$ . By the structure of prison strategies,  $\tau_1[p_1(l_2), \cdot]$  is constant. By NEWT,  $a_2 = br(a^*)$ . The only symmetric  $2 \times 2$  game, for which all these conditions hold, is a pure coordination game with a Pareto-dominant outcome.

Let us now show that in the class of coordination games with a Pareto-dominant outcome, equilibrium paths contain no more than a single period of miscoordination per cycle. Assume that there exist (at least) two periods  $k$  and  $l$  (where  $k < l$ ), such that  $a_1^k \neq a_2^k$  and  $a_1^l \neq a_2^l$ . By Lemma 1, period  $k$  already belongs to the cyclic phase. By the structure of prison strategies,  $p_1(l)$  and  $p_2(l)$  are not visited prior to  $k$ . By NEWT, the cyclic phase consists of strict best-replying. Therefore,  $e[p_1(l)] = e[p_2(l)] = 0$ , in contradiction with NEWT.

It remains to be shown that equilibrium paths contain no periods, in which exactly one player behaves non-myopically. The only symmetric  $2 \times 2$  game for which this is possible is the PD. Assume that there exists at least

one period, in which exactly one player plays  $C$  (we have already shown that there can be no periods, in which both players play  $C$ ).

First, observe that each player must play  $C$  at least once per cycle. Otherwise, there is at least one player who never plays  $C$  past a certain period. Therefore, there are a player  $j$  and a period  $k$ , such that  $a_j^k = C$  and  $a_{-j}^l = D$  for every  $l \geq k$ . By the payoff structure of PD,  $a_j^k \neq BR_j[p_{-j}(k)]$ , in contradiction to NEWT.

Now, consider the earliest period  $k$ , in which exactly one player, say player 1, plays  $C$ . We have shown that there must exist a period  $l > k$ , such that  $a_1^l = D$  and  $a_2^l = C$ . By Lemma 1,  $k$  already belongs to the cyclic phase, in which players adhere to best-replying. But by NEWT,  $e[p_1(l)] > 0$ . However, by the structure of prison strategies,  $p_1(l)$  is never visited prior to  $k$ , which means that  $e[p_1(l)] = 0$ , a contradiction. ■

Thus, non-myopic behavior in PD, Chicken or Battle-of-the-sexes cannot be sustained in NEWT with prison trigger strategies. Moreover, only sparse miscoordination can be sustained in coordination games with a Pareto-dominant outcome. This result illustrates the strength of the requirement that equilibrium strategies carry out both the best-replying and the supporting T-tests. Prison strategies are usually too tightly structured for this joint task.

This section has explored two subclasses of trigger strategies - “grim” and “prison” strategies. The former is totally unforgiving, whereas the latter is totally forgiving. The restrictiveness of both schemes largely derives from the fact that the duration of punishment is independent of the deviant player’s behavior during the punishment phase. This property makes it optimal for the player to behave myopically when he is being punished for a previous T-test. As demonstrated by Figure 2 (see  $s_2$ ), the ability to condition on players’ behavior during a punishment phase can be important for sustaining cooperation in a game such as PD. I do not know how the results in this section extend to arbitrary trigger strategies.

## 5 A Folk Theorem

The results gathered so far naturally raise the question of whether a folk theorem holds under NEWT. The answer turns out to be affirmative for arbitrary repeated  $2 \times 2$  (possibly asymmetric) games:

**Theorem 4** *Every individually rational payoff profile in a repeated  $2 \times 2$  game with discounting can be approximated by some NEWT, as the players' discount factor tends to 1.*

Here is the basic idea behind the construction for repeated Chicken, for example. The trick is to use the states in one of player 1's punishment phases, say, to execute the T-tests against player 2's strategy. (the output of punishment states can be “dove” as well as “hawk”, but of course there must be sufficiently many “hawk”-states amongst them, in order for the punishment to be effective). In contrast, the states in player 2's strategy that carry out the T-tests against player 1's strategy are “preliminary” states that are never revisited. Along the pre-cyclic phase, player 1 runs his required T-tests only after player 2 finished running his.

Because the game of Chicken satisfies the property that  $a = br[br(a)]$  for every  $a \in \{Hawk, Dove\}$ , the punishment and preliminary states can be endowed with constant transitions. This property is not satisfied by PD, for example, which makes the construction for PD is more elaborate.

This construction is made possible by an important property of non-zero-sum  $2 \times 2$  games:  $(a_j^*, br_{-j}(a_j^*))$  is a NE in the stage game, where  $a_j^*$  is the action that “minimaxes” player  $-j$ . Therefore, whenever player  $j$  “minimaxes” player  $-j$  and the latter responds myopically, no further T-tests are required. This is the very property that allows trigger strategies to generate an SPE folk theorem in  $2 \times 2$  games. I do not know whether the folk theorem would extend to larger games, which generally lack this property.

## 6 Games with Non-Simultaneous Moves

Although Definition 1 is stated for repeated games with simultaneous moves, the basic idea underlying NEWT is extendible to a broader class of infinite-horizon games, including non-simultaneous moves or non-time-separable preferences. For example, the notion of myopic behavior is meaningless in bargaining interactions, which usually do not involve periodic flow of payoffs. Nevertheless, the idea of testing threats that affect optimal response is equally pertinent to bargaining interactions.

Rather than presenting a “complete” generalization of NEWT, let us study a special case. Consider an infinite-horizon, multi-period two-person game. In each period  $t = 0, 1, 2, \dots$ , player 1 moves first and player 2 moves



second. Each player  $j$  at his turn faces an action set  $A_j$  of size two. Preferences are stationary, but they need not be time-separable. Further extensions (number of players, number of actions, preferences with history-dependent evaluation of continuation paths) are straightforward.

Let us continue to represent player  $j$ 's pure strategy by a finite automaton  $(Q_j, q_j^0, f_j, \tau_j)$ , with the convention that  $q_j^0$  is the player's state at the beginning of the game and  $\tau_j$  carries him to the next state, given the current state  $q$  and player  $-j$ 's latest action. In particular,  $p_2(0)$  is not  $q_2^0$ , but  $\tau_2(q_2^0, a_1^0)$ . By the stationarity of preferences, player  $-j$ 's evaluation of continuation paths depends only on the current state of  $s_j$ . Thus, there is a well-defined  $BR$  function. Using a slightly fussier notation than in Section 2, let  $BR_{-j}(s_j \mid q)$  stand for the set of actions prescribed for player  $-j$  by best-replying to  $s_j$ , conditional on player  $j$  being in state  $q$  (again, assume for simplicity that  $BR_{-j}(s_j \mid q)$  is a singleton).

For every  $q \in Q$  in  $s_j = (Q, q^0, f, \tau)$ , define  $r_q(s_j) = (Q, q^0, f, \tau')$ , which is identical to  $s_j$ , except that  $\tau'(q, a) = \tau(q, BR_{-j}(s_j \mid q))$  for every  $a \in A_{-j}$ . The automaton  $r_q(s)$  is a simplification of  $s$ . According to  $s$ , the best-reply action against  $q$  leads to a certain continuation. According to  $r_q(s)$ , this is the continuation **regardless** of player  $j$ 's action against  $q$ . When  $\tau(q, \cdot)$  is constant,  $r_q(s) = s$ .

**Definition 4**  $(s_1, s_2)$  is a NEWT if for every player  $j = 1, 2$ :

1.  $a_j^k = BR_j[s_{-j} \mid p_{-j}(k)]$  for every sufficiently large period  $k$  along  $z(s_1, s_2)$ .
2.  $e(q) > 0$  if and only if  $BR_j(s_{-j} \mid q) \neq BR_j[r_q(s_{-j}) \mid q]$ , for every state  $q \in Q_{-j}$  that is visited along  $z(s_1, s_2)$ .

The modified second condition says that if  $a \in A$  is the best-reply action against  $q$  only because of the perceived threat to punish  $-a$ , then the player must test the threat first, by playing  $-a$  against  $q$ . In repeated games with discounting, this is equivalent to Definition 1. The remainder of this section applies this modified definition of NEWT to a couple of games.

## 6.1 The Sequential-Move Prisoner's Dilemma

In Section 4, we saw that mutual cooperation in repeated PD cannot be sustained in NEWT with prison strategies. Let us consider a sequential-move version of the PD, in which player 1 (2) is the stage-game leader (follower). The following profile of prison strategies is a NEWT:

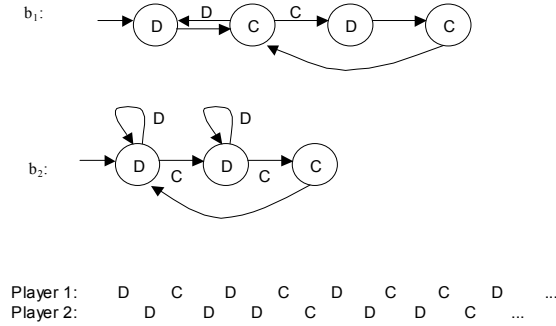


Figure 5

Where lies the difference between the simultaneous- and sequential-move cases? Observe that at period 4, both players play  $C$ , which is a non-myopic action and therefore requires T-tests. However, player 1's  $C$  is a best-reply action against  $p_2(3)$ , not  $p_2(4)$ . This means that player 1's (2's) T-test is supposed to justify his cooperativeness against player 2's (1's) defective (co-operative) behavior. This asymmetry does not hold in the simultaneous-move case. This example illustrates that the class of NEWT in a repeated game is sensitive to whether moves are simultaneous or sequential.

## 6.2 A Repeated Ultimatum Game

Consider an infinite repetition (with discounting) of the following extensive-form game:

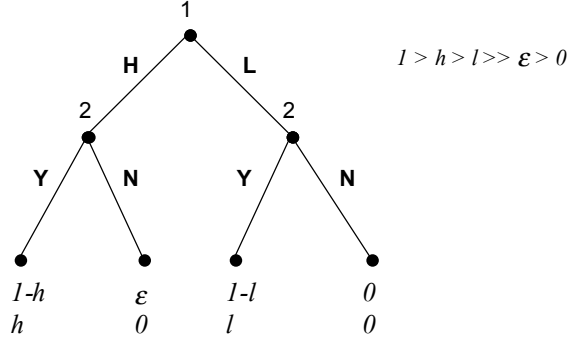


Figure 6

This is a two-action version of the Ultimatum Game ( $H$  and  $L$  stand for a generous offer and a mean offer, whereas  $Y$  and  $N$  stand for accepting and rejecting an offer). The only non-standard feature of the payoff structure is the proposer's small positive payoff when the responder declines a generous offer, which I introduce as a tie-breaker. Given that the proposer's offer is rejected, he is slightly better off being generous.

The repeated game has a trivial NEWT, in which the proposer always makes a mean offer, which the responder always accepts. In this subsection, we will study play paths, in which player 1 keeps making a generous offer that is accepted by player 2, from some period  $k^*$  onwards. As in Section 3.2, let us restrict attention to strategies that display constructive reciprocity (the minimax actions are  $a_1^* = L$  and  $a_2^* = N$ ). This restriction entails no loss of generality under NE. Whenever persistently generous behavior is NE-sustained at all, it is sustainable by NE with CR strategies. The following result demonstrates that under NEWT, restricting attention to CR strategies carries strong implications:

**Theorem 5** *Suppose that  $(s_1, s_2)$  is a NEWT with CR strategies, such that  $a_1^k = H$  and  $a_2^k = Y$  for every period  $k > k^*$  along  $z(s_1, s_2)$ . Let the players' discount factor tend to 1. Then,  $h < \frac{2+l}{3}$ .*

Thus, the combination of NEWT and CR strategies in the repeated Ultimatum game carries a strong **payoff restriction**. In order for constructive reciprocity to be consistent with sustained generosity, the gap between  $h$  and  $l$  cannot be too large. What drives this result is that player 2 does not need to back-up his cyclic behavior with T-tests (contrary to repeated PD, for example), because  $Y = br_2(H)$ .

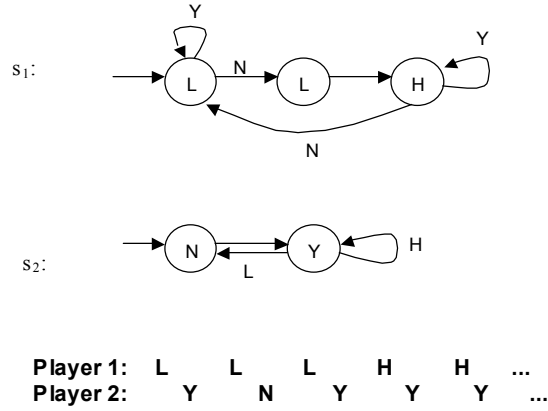


Figure 7

Figure 7 represents a profile of CR strategies, which constitutes a NEWT as long as  $h < \frac{1+l}{2}$ . The play pattern is as follows: player 1 initially makes mean offers; he gets away with the first but encounters rejection of the second; then, he makes one last mean offer before switching to generous behavior. The proposer's switch to generous offers following an acceptance, rather than a rejection of a mean offer, is somewhat counter-intuitive. The reason is that the “lesson” the proposer draws from his T-tests is that he cannot get away with **two** consecutive mean offers, but he can get away with one.

## 7 Discussion

The concept of NEWT is based on a new notion of “credible threats” in infinite-horizon games: threats are “credible” if they were actually tested and

realized in the past. The particular way in which this idea is incorporated into the definition of NEWT raises questions of interpretation, which I would now like to discuss.

## 7.1 Interpreting the Equilibrium Condition

Following the convention of Aumann (1987) and Rubinstein (1991), player  $j$ 's equilibrium strategy  $s_j$  is interpreted as a **belief** held by his opponent  $-j$ . However, when the motivation for threat testing is genuine strategic uncertainty, it seems implausible to interpret  $s_j$  as the belief that player  $-j$  holds at the beginning of the game. Instead, equilibrium strategies should be interpreted as the beliefs that players come to hold **by the time they enter the cyclic phase**. The assumption underlying NEWT is that in order for the players to have arrived at those beliefs, they must have tested the relevant threats in the pre-cyclic phase. However, the concept makes no explicit presumption concerning the players' beliefs at the beginning of the game.

Throughout the paper, there has been no attempt to rationalize the model of behavior underlying NEWT. In this respect, the paper shares a similar approach with recent works in the bounded rationality literature, which formulate equilibrium concepts on the basis of procedurally rational behavior: McKelvey and Palfrey (1995), Osborne and Rubinstein (1998), Spiegler (2000), Eliaz (2001), Jehiel (2001), among others.

Of course, it is of interest to know whether there are conditions, in which NEWT is consistent with standard Bayesian rationality. For example, consider an arbitrarily patient, Bayesian rational player, who faces a "forgiving" strategy (i.e., all the punishments are reversible, as in prison strategies). For each of the opponent's threats, he places a positive probability on the event that the threat does not exist. Such a player will conform to our model. (such that  $e(q) \leq 1$  for every  $q$ .) In this case, we can say that NEWT captures the last stage of a process of learning to play a repeated-game NE, in addition to the NE itself.

T-tests cannot be universally rationalized. E.g., a Bayesian rational player will not want to test a "grim" threat if he is sufficiently patient. In general, testing irreversible punishments is harder to rationalize. However, irreversible punishments are less interesting to begin with in our framework, as Remark 1 illustrates. Long-run non-myopic behavior, which is sustained by an irreversible punishment, will not survive the T-tests and is therefore

inconsistent with NEWT.

This entire problem of interpreting equilibrium strategies vanishes when T-tests are driven by self-control or justifiability considerations, rather than by strategic uncertainty. In this case, the equilibrium condition of NEWT is interpreted in the same way as in NE. Players have perfect forecasts of each other's strategy, but their individual decision procedure is different - rather than choosing a course of action that maximizes utility, they choose a course of action they can justify, for example.<sup>5</sup>

## 7.2 The Finite Automata Representation

The threat testing procedure is formalized using finite automata representation of repeated-game strategies. Testing a threat corresponds to traveling through the transition  $\tau[q, -BR(q)]$  from the state  $q$ . Two histories in which the opponent is in the same state (different states) are (not) equivalent for threat-testing purposes.

This is the most conservative way of distinguishing between different threats (see Kalai and Stanford (1988)). Due to the stationarity of the infinitely repeated game, the set of available continuation paths for player  $j$ , given  $s_{-j}$ , is completely determined by the state in which  $s_{-j}$  is. Two histories in which  $s_{-j}$  is in the same state are not only payoff-equivalent for player  $j$ , but they are also equivalent in terms of his set of feasible continuation paths.

The converse is not always true because the finite automata representation may contain "redundancies". For example, the following automaton represents the same strategy as Figure 1:  $Q = \{q_1, q_2, r\}$ ,  $f(q_1) = f(q_2) = C$ ,  $f(r) = D$ ,  $\tau(q_1, D) = \tau(q_2, D) = r$ ,  $\tau(q_1, C) = q_2$ ,  $\tau(q_2, C) = q_1$ ,  $\tau(r, C) = r$  and  $\tau(r, D) = q_1$ . Thus, one of the two  $C$ -states in this representation is redundant. Nevertheless, NEWT treats the two threats associated with the two  $C$ -states as two different threats, which require at least one T-test each. Therefore, NEWT the minimal overall number of T-tests that are required against this representation is higher than against the representation given by Figure 1.

It follows that the finite automata representation is a non-trivial component of the definition of NEWT - different representations of the same

---

<sup>5</sup>It could be argued that justifiability is not a decision procedure, but a particular preference. Under this interpretation, NEWT is a reduced-form model. The reader is referred to Spiegler (2000) for an elaborate discussion.

strategies give rise to different equilibrium behavior. **However, none of the results in this paper depends on this feature.**

It is interesting to compare this paper to the literature on complexity considerations in repeated games, pioneered by Rubinstein (1986) and Neyman (1985), which also relies heavily on the finite automata formalism. (although the present paper employs the formalism as convenient language for expressing the idea of systematic threat testing, not to capture complexity considerations.)

Rubinstein (1986) and Abreu and Rubinstein (1988) study NE in repeated games, in which players also prefer a smaller number of states in their automaton. Banks and Sundaram (1990) assume that players also prefer a smaller number of transitions. An immediate consequence of these assumptions is that in equilibrium, all states are visited (in Banks and Sundaram (1990), all transitions are traveled as well). These properties are also satisfied in NEWT, whenever equilibrium strategies satisfy that every state is reachable from every other state.

The crucial property that distinguishes NEWT from these models is the following: if  $BR(q) \neq br[f(q)]$  and  $q$  is visited in  $z(s_1, s_2)$ , then both  $BR(q)$  and  $br[f(q)]$  are played against  $q$  along  $z(s_1, s_2)$ . In other words, there cannot be one-to-one correspondence between the two players' actions in any NEWT that contains non-myopic behavior. In contrast, such one-to-one correspondence is a central property of the above models.

### 7.3 Other Testing Procedures

T-tests are not the only way of checking contingencies in repeated games. E.g., even when a player in the repeated PD expects his opponent's strategy to be "always defect", he may occasionally try cooperative behavior, in the hope that this will trigger reciprocal cooperation by the opponent. This type of experimentation is ruled out by NEWT. Allowing for arbitrary experimentation in the pre-cyclic phase, in addition to T-tests, can be achieved simply by removing the "only if" in Condition 2 of Definition 1.

However, there is a difference in principle between T-tests and such alternative procedures. The objective of T-tests is to verify the contingencies that are postulated by the player's belief, and those are **finite** in number. in contrast, the latter type of experiments is designed to discover new contingencies that are not postulated by the belief, and those are **infinite** in number. This is analogous to the difference between sample-error tests ( $t$ -tests) and

specification-error tests in classical statistical inference - the former is normally perceived as a more rudimentary and standard procedure. Clearly, the classical statistician sometimes carries out both procedures. Nevertheless, if we wanted to describe her normal behavior, focusing on  $t$ -tests and ignoring specification-error tests would be a good “first approximation”.

## 8 References

1. Abreu D. and A. Rubinstein (1988). “The Structure of Nash Equilibrium in Repeated Games with Finite Automata.” *Econometrica* **56**, 1259-1282.
2. Aumann R. J. (1987). “Correlated Equilibrium as an Expression of Bayesian Rationality.” *Econometrica* **55**, 1-18.
3. Banks J. and R. Sundaram (1990). “Repeated Games, Finite Automata and Complexity.” *Games and Economic Behavior* **2**, 97-117.
4. Eliaz, K. (2001). “Nash Equilibrium when Players Account for the Complexity of their Forecasts.” Mimeo.
5. Jehiel P. (2001). “Analogy-Based Expectation Equilibrium.” Mimeo.
6. Kalai E. and W. Stanford (1988). “Finite Rationality and Interpersonal Complexity in Repeated Games.” *Econometrica* **56**, 387-410.
7. McKelvey, R. D. and T. R. Palfrey (1995). “Quantal Response Equilibria for Normal Form Games.” *Games and Economic Behavior* **10**, 6-38.
8. Neyman A. (1985). “Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner’s Dilemma.” *Economic Letters* **19**, 227-229.
9. Osborne M. and A. Rubinstein (1994). *A Course in Game Theory*. MIT Press.
10. Osborne and Rubinstein (1998). “Games with Procedurally Rational Players.” *American Economic Review* **88**, 834-847.
11. Rubinstein A. (1986). “Finite Automata Play the Repeated Prisoner’s Dilemma.” *Journal of Economic Theory*



12. Rubinstein A. (1991). “Comments on the Interpretation of Game Theory.” *Econometrica* **59**, 909-924.
13. Selten R., M. Mitzkewitz and G. R. Uhlich (1997). “Duopoly Strategies Programmed by Experienced Players.” *Econometrica* **65**, 517-555.
14. Spiegler R. (2000). “Reason-Based Choice and Justifiability in Extensive-Form Games.” Mimeo.
15. Tetlock, P.E., and R. Boettger (1989). “Accountability: A Social Magnifier of the Dilution Effect.” *Journal of Personality and Social Psychology* 57, 388-298.
16. Tetlock, P.E., and R. Boettger (1994). “Accountability Amplifies the Status Quo Effect when Change Creates Victims.” *Journal of Behavioral Decision Making* 7, 1-23.

## 9 Appendix: Proofs

### 9.1 Theorem 1

Let us first show that  $k^C \geq 6$ . Since  $C \neq br(C)$ , NEWT implies that for both  $j = 1, 2$ ,  $C = BR_j[p_{-j}(k^C)]$  and there exists at least one period  $k_j < k^C$ , such that  $p_j(k_j) = p_j(k^C)$  and  $a_{-j}^{k_j} = D$ . Moreover,  $k_j \neq k^C - 1$ ; otherwise, the payoff structure of PD implies  $BR_{-j}[p_j(k^C)] = D$ , a contradiction. Also, since  $C \neq br(D)$ , NEWT implies that for both  $j = 1, 2$ ,  $C = BR_j[p_{-j}(k_{-j})]$  and there exists at least one period  $l_j < k_j$ , such that  $p_j(l_j) = p_j(k_j)$  and  $a_{-j}^{l_j} = D$ . It follows that  $k^C \geq 5$ .

Suppose that  $k^C = 5$ . Then, the play path must be as follows (assume  $k_1 < k_2$ , without loss of generality):

Player 1:	$D$	$C$	$D$	$D$	$C$	$C$	...
Player 2:	$D$	$D$	$C$	$D$	$C$	$C$	...

Observe that  $\tau_1[p_1(1), D] = p_1(k^C)$ . Therefore,  $BR_2[p_1(1)] = D$ , a contradiction.

Let us now turn to comparative statics. The reader can verify that for any PD payoffs, there exists  $M \geq 1$ , such that the following strategy profile

constitutes a NEWT for every  $N \geq M$  (the rectangle represents  $M - 1$  consecutive  $D$ -states with constant transitions):

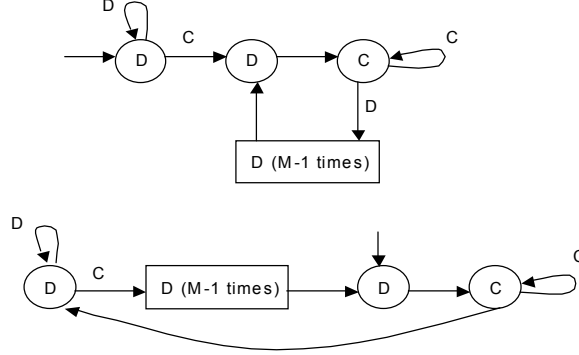


Figure 8

It remains to be shown that the build-up phase can never be shorter than the one given in Figure 8. Let  $l^*$  be the latest period  $l$  for which  $a_1^l \neq a_2^l$ , in some NEWT with minimal cooperation build-up. As we have already observed,  $l^* < k^C - 1$ . Suppose, without loss of generality, that  $a_1^{l^*} = C$ . Then,  $p_1(l^*)$  must be visited in the cycle. Otherwise, there would be a NEWT with a shorter build-up phase, which differs from the NEWT in question only in that  $\tau_j[p_j(l^* - 1), a_{-j}^{l^*-1}] = p_j(k^C)$  for every  $j = 1, 2$ , thereby contradicting the minimality assumption. Now, in order for  $BR_2[p_1(l^*)] = C$  to hold, it must be that  $k^C - l^* - 1 \geq M$ , by the definition of  $M$ . We have already shown that  $l^* \geq 4$  (while proving that  $k^C \geq 6$ ). It follows that  $k^C \geq M + 5$ . Finally, it is easy to verify that  $M$  is increasing in the players' discount factor and decreasing in  $u(C, C)$ .

## 9.2 Lemma 1

Suppose that  $a_j^k \neq br_j(a_{-j}^k)$  for **both**  $j = 1, 2$ . For each  $j \in \{1, 2\}$ , there is a finite number  $N_j$  of states  $q \in Q_j$  satisfying  $f_j(q) = a_j^k$  and  $BR_{-j}(q) \neq br_{-j}(a_j^k)$ , that are visited along  $z(s_1, s_2)$ . Moreover, the structure of prison strategies implies that these states are visited in a fixed order  $q^1, \dots, q^{N_j}, q^1, \dots$ ,

such that  $q^{(n+1) \bmod N_j}$  is visited only after  $BR_{-j}(q^{n \bmod N_j})$  is played against  $q^{n \bmod N_j}$ . Therefore, there must exist a period  $k' > k$  such that  $p_j(k') = p_j(k)$  for both  $j \in \{1, 2\}$ , which means that  $k$  belongs to the cyclic phase of  $z(s_1, s_2)$ .

Now, suppose that  $a_j^k \neq br_j(a_{-j}^k)$  for exactly **one** player  $j$ . The only symmetric  $2 \times 2$  game, for which this is possible, is the PD. Without loss of generality, let  $a_1^k = C$  and  $a_2^k = D$ . By NEWT,  $a_1^k = BR_1[p_2(k)]$ . Clearly,  $s_1$  contains a finite number of  $C$ -states. Similarly,  $s_2$  contains a finite number  $M$  of  $D$ -states  $q \in Q_2$  satisfying  $BR_2(q) = C$ . By the structure of prison strategies, these  $D$ -states are visited in a fixed order  $q^1, \dots, q^M, q^1, \dots$ , such that  $q^{(n+1) \bmod M}$  is visited only after  $BR_{-j}(q^{n \bmod M}) = C$  is played against  $q^{n \bmod M}$ . Therefore, there must exist a period  $k' > k$  such that  $p_j(k') = p_j(k)$  for both  $j \in \{1, 2\}$  - i.e.,  $k$  belongs to the cyclic phase of  $z(s_1, s_2)$ .

### 9.3 Theorem 4

Consider a cycle of actions that yields individually rational payoffs. Construct a strategy profile  $(s_1, s_2)$ , which induces this cycle in its cyclic phase. Let  $(r_1^1, \dots, r_1^m)$  and  $(r_2^1, \dots, r_2^n)$  be the states in  $s_1$  and  $s_2$ , respectively, that are visited in the cyclic phase and satisfy  $BR_{-j}(r) \neq br_{-j}[f_j(r)]$ . The states are enumerated according to their order of appearance in the cycle. Denote by  $a_j^*$  the action for player  $j$  that minimaxes player  $-j$ . Along  $z(s_1, s_2)$ , let  $k$  be an arbitrarily large period in the cyclic phase, for which  $p_2(k) = r_2^1$ . Denote  $p_1(k) = p^*$ .

For simplicity, let us first consider  $2 \times 2$  games, for which  $(a_1, br_2(a_1))$  is a stage-game NE for every action  $a_1$  (Chicken and Battle-of-the-sexes fall into this category).

Player 2's strategy  $s_2$  is constructed as follows:

1. For every  $k = 1, \dots, n$ ,  $\tau_2[r_2^k, -BR_1(r_2^k)]$  begins a sequence of  $N^k$  constant-transition states, whose output is  $a_2^*$ . This sequence in turn leads to the state  $r_2^{k+1 \bmod n}$ . The numbers  $(N^k)_{k=1, \dots, n}$  are chosen to be big enough, such that  $BR_1(r_2^k) \neq br_1[f_2(r_2^k)]$ .
2.  $f_2(q_2^0) = br_2[f_1(r_1^1)]$  and  $\tau_2(q_2^0, \cdot)$  is constant.
3.  $q_2^0$  leads to a sequence of  $M^1$  constant-transition states, whose output is  $br_2(a_1^*)$ , where  $M^1$  is arbitrarily large.

4. This sequence of states leads to a constant-transition state  $p_2^1$ , such that  $f_2(p_2^1) = br_2[f_1(r_1^2)]$ .
5.  $p_2^1$  leads to a sequence of  $M^2$  constant-transition states, whose output is  $br_2(a_1^*)$ , where  $M^2$  is arbitrarily large.
6. This construction proceeds along the same lines, until we hit  $p_2^m \equiv r_2^1$ .

Player 1's strategy  $s_1$  is constructed as follows:

1.  $q_1^0 = r_1^1$ .
2. For every  $k = 1, \dots, m$ ,  $\tau_1[r_1^k, -BR_2(r_1^k)]$  begins a sequence of  $M^k$  constant-transition states, whose output is  $a_1^*$ , where  $M^k$  is as specified in the construction of player 2's strategy). Furthermore:
  - (a) For every  $k = 1, \dots, m-1$ , the  $k^{\text{th}}$  sequence leads to  $r_1^{k+1}$ .
  - (b) The  $m^{\text{th}}$  sequence leads to  $n$  concatenated sequences of constant-transition states. For every  $h = 1, \dots, n$ , the  $h^{\text{th}}$  sequence consists of one state, whose output is  $br_1[f_2(r_2^k)]$ , followed by  $N^h$  states, whose output is  $br_1(a_2^*)$ . The  $n^{\text{th}}$  sequence leads to  $p^*$ .

It is straightforward to verify that this strategy profile constitutes a NEWT. The cycle begins in the earliest period  $k$ , for which  $p_1(k) = p^*$  and  $p_2(k) = r_2^1$ .

A modification is needed whenever  $f_j(r_j^k) \neq br_j[br_{-j}(f_j(r_j^k))]$  for some  $r_j^k$  (e.g., when  $f_j(r_j^k) = C$  in the PD). By the construction, there is a unique period  $l$  in the pre-cyclic phase, for which  $p_j(l) = r_j^k$ . Also, by construction,  $f_{-j}[p_{-j}(l)] = br_{-j}(r_j^k)$  and  $\tau_{-j}[p_{-j}(l), \cdot]$  is constant.

The modification of  $s_{-j}$  is as follows. Instead of  $\tau_{-j}[p_{-j}(l), f_j(r_j^k)] = \tau_{-j}[p_{-j}(l), f_j(r_j^k)]$ , let  $\tau_{-j}[p_{-j}(l), f_j(r_j^k)]$  begin a sequence of  $K$  constant-transition states, whose output is  $a_2^*$ . This sequence leads back to  $p_{-j}(l)$  and  $K$  has to be sufficiently high to ensure that  $f_j(r_j^k) \neq BR_j[p_{-j}(l)]$ . The modification of  $s_j$  is as follows. By the construction,  $r_j^k$  has a unique state  $q$  that is not visited in the cycle and satisfies  $\tau(q, \cdot) = r_j^k$ . Between  $q$  and  $r_j^k$ , insert  $K$  constant-transition states, whose output is  $br_j(a_{-j}^*)$ , where  $K$  is as defined above. it is straightforward to verify that this modification results in a NEWT.

## 9.4 Theorem 5

By constructive reciprocity, if player 2 always plays  $Y$  against an  $H$ -state, he guarantees the maximal continuation payoff that is feasible for him in the game. Therefore,  $e(q') = 0$  for every  $H$ -state  $q' \in Q_1$ . Thus,  $k^*$  is the earliest period, in which player 1 plays  $H$ .

On the other hand,  $\tau_2(q, H) = Y$  and  $BR_1[s_2 \mid q] = H$  for every state  $q \in Q_2$  that is visited in the cycle (at least from period  $k^* + 1$  onwards). Therefore,  $BR_1[r_p(s_2) \mid q] = L$  and so,  $e(q) > 0$  and  $a_1^1 = L$ . By constructive reciprocity,  $\tau_2(q, L) = N$ . It follows that there exists a period  $k < k^*$ , such that  $p_2(k-1) = q$ ,  $a_1^k = L$  and  $a_2^k = N$ . By NEWT,  $a_2^k = BR[s_1 \mid p_1(k)]$  and  $e[p_1(k)] > 0$ . Therefore,  $a_2^1 = Y$ ,  $a_1^2 = L$ , and by constructive reciprocity,  $a_2^2 = N$ .

There are two cases to consider:

1.  $a_1^3 = L$ . Suppose that  $a_2^3 = N$ . Then,  $BR_1(s_2 \mid p) = L$  and  $BR_1[r_p(s_2) \mid p] = H$ . NEWT requires that  $e(p) > 0$ , which is not satisfied along the play path, a contradiction. It follows that  $a_2^3 = Y$ . By constructive reciprocity and since players are arbitrarily patient, player 1 can guarantee a discounted payoff of  $1 - h$  against  $p_2(3)$ , which is the same as the payoff he can guarantee against  $p_2(1)$ . Therefore, to ensure that  $BR(s_2 \mid q) = H$  for any  $Y$ -state  $q \in Q_2$  that is visited in the cyclic phase, stage-game payoffs must satisfy  $h < \frac{1+l}{2}$ .
2.  $a_1^3 = H$ . Then,  $k^* = 3$  and so,  $p_2(1)$  must be visited in the cyclic phase. Denote  $p_2(2) = p$ . Suppose that  $BR_1(s_2 \mid p) = H$ . Since  $BR_1[r_p(s_2) \mid p] = L$ , NEWT requires that  $e(p) > 0$ , which is not satisfied along the play path, a contradiction. Therefore,  $BR_1(s_2 \mid p) = L$ . If  $f_2[\tau_2(p, L)] = Y$ , then by constructive reciprocity, player 1 can guarantee a higher payoff by playing  $L$  against  $p$  than by playing  $H$ , thus contradicting the fact that  $L$  is the best-reply action against  $p$ . It follows that  $f_2[\tau_2(p, L)] = N$ . To ensure that  $BR_1[s_1 \mid p_2(1)] = H$  and  $BR_1[s_2 \mid p_2(2)] = L$ , stage-game payoffs must satisfy  $h < \frac{2+l}{3}$ .